

TESTING THE NORMALITY OF THE GRAVITATIONAL WAVE DATA WITH A LOW COST RECURSIVE ESTIMATE OF THE KURTOSIS

E. Chassande-Mottin

ILGA-Virgo (CNRS – UMR6162), Observatoire de la Côte d’Azur
BP 4229 F-06304 Nice Cedex 4 France

ABSTRACT

We propose a monitoring indicator of the normality of the output of a gravitational wave detector. This indicator is based on the estimation of the kurtosis (i.e., the 4th order statistical moment normalized by the variance squared) of the data selected in a time sliding window. We show how a low cost (because recursive) implementation of such estimation is possible and we illustrate the validity of the presented approach with a few examples using simulated random noises.

Four large-scale detectors [1] of gravitational waves (GWs) are on the point to take their first scientific data. They all rely on the principle used in the Michelson experiment: measure the relative length $\delta L/L$ of the two perpendicular arms (each formed by two suspended masses) of the detector. The goal is to reach a measurement sensitivity (i.e., decrease the noise level) such that the small changes of δL caused by the GWs emitted from astrophysical sources (such as the coalescing binaries of neutron stars or black holes) can be detected when they pass through the instrument.

Since only a small number of such events are likely to be observed, the problem consists from the data analysis viewpoint, in looking for rare transients appearing in the detector output. For the coalescing binaries mentioned above, this will be done by implementing a bank of matched filters: each filter correlate the data with the expected GW emitted from a binary and we repeat this operation for a large number of possible targets. Assuming the template waveform are reliable, the matched filtering approach can be shown to be optimal in the Neymann-Pearson sense provided that the noise is Gaussian. Therefore, it is crucial to monitor the noise Gaussianity and locate any departures from this nominal hypothesis.

There are many ways for the noise to departs from Gaussianity, however not all of them are relevant for the problem of GW detection. One of the possibilities is to have a noise probability density function (PDF) with heavier tails than the Gaussian bell curve. This particular discrepancy is a problem because it causes a increase of the false detection rate (i.e., the large values in the tails make the matched filter triggers more often).

The kurtosis [2] is a well-known measurement of the decay rate of the PDF in the tails. This motivates us to use

The author thanks the European Gravitational Observatory for its hospitality.

the kurtosis as an index measuring normality. We define the mean of the signal $x(t)$ by $\mu_1(t) \equiv \mathbb{E}[x(t)]$ where $\mathbb{E}[\cdot]$ is the expectation operator. The central moments [2] of $x(t)$ are given by $\mu_r(t) \equiv \mathbb{E}[(x(t) - \mu_1(t))^r]$ where $r \geq 2$ is the order. The kurtosis is defined by the following ratio $\kappa_4(t) \equiv \mu_4(t)/(\mu_2(t))^2$.

Because the analysis must be done in real-time, the normality test should not be computationally expensive. We propose here an efficient (because recursive) implementation of the kurtosis estimation in a short-term and sliding observation window. Another recursive estimator of the kurtosis was proposed in [3] and required to have zero-mean signals. The presented approach here works also with non-centered signals.

The outline of the paper is as follows. We choose a simple mathematical structure for the short-term and recursive estimation of the central statistical moments. In Sect. 1, we identify in the selected family which estimators of μ_1 , μ_2 and μ_4 have a vanishing bias. We then show in Sect. 2 that an adequate Taylor approximation of the ratio of the unbiased estimates of μ_4 and μ_2 obtained previously yields the recursive estimate of κ_4 and we detail its computation algorithm. Finally, we apply in Sect. 3 the proposed estimator to a few illustrative cases and we explain how it is used in practice for the monitoring of the normality of the GW detector output.

1. RECURSIVE ESTIMATES WITH VANISHING BIAS

1.1. Recursive estimate of the mean and variance

We assume that the signal $x(t)$, $t \in \mathbb{Z}$ (using a unit sampling rate) is locally stationary (i.e., its statistical moments do not change during a finite time period δt_{stat}) and ergodic (i.e., its statistical moments can be estimated from its samples). The mean of $x(t)$ can be estimated with the following weighted average of the data selected by the window $w_1(t)$

$$\hat{\mu}_1(t) \equiv C_1 \sum_{n=0}^t w_1(t-n) x(n), \quad (1)$$

where the duration of $w_1(t)$ is smaller than δt_{stat} . If the window in use is of exponential type, i.e. $w_1(t) \equiv a_1^t$, this estimator can be equivalently calculated recursively with

$$\hat{\mu}_1(t) = a_1 \hat{\mu}_1(t-1) + C_1 x(t). \quad (2)$$

The problem is to find the constants $a_1 > 0$ and C_1 such that $\hat{\mu}_1(t)$ is asymptotically unbiased i.e., $\mathbb{E}[\hat{\mu}_1(t)] = \mu_1$ when $t \rightarrow +\infty$. The bias can be calculated directly from the definition of the estimator, yielding

$$\mathbb{E}[\hat{\mu}_1(t)] = C_1 \mu_1 \frac{1 - a_1^{t+1}}{1 - a_1} \rightarrow \frac{C_1}{1 - a_1} \mu_1 \quad \text{when } t \rightarrow +\infty$$

and assuming $a_1 < 1$. We conclude that $\hat{\mu}_1(t)$ is an unbiased estimator of the mean if $C_1 = 1 - a_1$.

The same method can be applied to find recursive and unbiased estimators of the higher order moments, based on the following choice of expression :

$$\hat{\mu}_r(t) \equiv C_r \sum_{n=0}^t w_r(t-n) (x(n) - \hat{\mu}_1(n-1))^r, \quad (3)$$

where $r \geq 2$. Similarly to the mean, a recursive implementation is possible if we choose $w_r(t) \equiv a_r^t$. We restrict to the interval $0 < a_r < 1$.

When $r = 2$, we essentially average the squared differences to the estimated mean in a sliding time window defined by $w_2(t)$. We evaluate the bias of $\hat{\mu}_2(t)$ by taking the expectation of (3). We first evaluate :

$$\begin{aligned} \mathbb{E}[(x(n) - \hat{\mu}_1(n-1))^2] &= \mu_1^2 + \mu_2 - 2C_1 \mu_1^2 \frac{1 - a_1^n}{1 - a_1} \\ &\quad + C_1^2 \left(\mu_2 \frac{1 - a_1^{2n}}{1 - a_1^2} + \mu_1^2 \left(\frac{1 - a_1^n}{1 - a_1} \right)^2 \right), \end{aligned}$$

and setting $C_1 = 1 - a_1$ (i.e., set the bias of $\hat{\mu}_1$ to zero), the summation leads to :

$$\begin{aligned} \mathbb{E}[\hat{\mu}_2(t)] &= \frac{2C_2 \mu_2}{1 + a_1} \frac{1 - a_2^{t+1}}{1 - a_2} + C_2 \left(\mu_1^2 - \frac{1 - a_1}{1 + a_1} \mu_2 \right) \\ &\quad \times \frac{a_2^{t+1} - a_1^{2(t+1)}}{a_2 - a_1^2} \rightarrow \frac{2C_2}{(1 + a_1)(1 - a_2)} \mu_2, \end{aligned}$$

when $t \rightarrow +\infty$ provided that a_1 and $a_2 < 1$. Consequently, the bias of $\hat{\mu}_2(t)$ is zero when $C_2 = (1 + a_1)(1 - a_2)/2$.

1.2. Extension to the 4th order

We proceed to the fourth order as previously, starting from the definition (3) with $r = 4$. The evaluation of the bias of $\hat{\mu}_4(t)$ requires the calculation of

$$\begin{aligned} \mathbb{E}[(x(n) - \hat{\mu}_1(n-1))^4] &= (1 + f_4) \mu_4 \\ &\quad + 4(1 - f_1 - f_3 + f_1 f_3) \mu_1 \mu_3 + 3(2f_2 + f_2^2 - f_4) \mu_2^2 \\ &\quad + 6(1 - 2f_1 + f_2 + f_1^2 - 2f_1 f_2 + f_1^2 f_2) \mu_1^2 \mu_2 \\ &\quad + (1 - 4f_1 + 6f_1^2 - 4f_1^3 + f_4) \mu_1^4, \end{aligned}$$

where we defined $f_k \equiv C_1^k (1 - a_1^{kn}) / (1 - a_1^k)$.

Setting $C_1 = 1 - a_1$ so that the estimate of the mean is not biased, and making the summation for $n = 0 \dots t$, we get

$$\mathbb{E}[\hat{\mu}_4(t)] = \frac{C_4}{1 - a_4} (k_4 \mu_4 + k_2 \mu_2^2) + r(t), \quad (4)$$

where $r(t)$ is a complicated sum of integer powers of terms of the form $K a_1^{(t+1)}$ and $K a_4^{(t+1)}$ where $K \in \mathbb{R}$. The constants can be expressed as

$$k_4 = \frac{2(1 - a_1 + 2a_1^2)}{(1 + a_1)(1 + a_1^2)} \quad k_2 = \frac{6(1 - a_1)(1 + a_1 + 2a_1^2)}{(1 + a_1)^2(1 + a_1^2)}.$$

It is interesting to note that the first term in (4) does not depend on μ_3 .

If a_1 and $a_4 < 1$, the function $r(t)$ goes to 0 when t tends to $+\infty$ so that the expectation of $\hat{\mu}_4(t)$ depends in a simple manner of μ_4 (the objective value) and μ_2^2 . Assuming that μ_2 is known, we can set the bias of $\hat{\mu}_4(t)$ to a simple constant offset if we choose the coefficient of μ_4 in (4) equal to 1, i.e. we set $C_4 = (1 - a_4)/k_4$ in the following.

2. TAYLOR APPROXIMATION OF THE KURTOSIS ESTIMATOR

The results of the previous Section motivate us to propose the following estimator $\hat{\kappa}_4(t)$ of the kurtosis, obtained by dividing (4) by μ_2^2 and replacing the variance by its recursive estimate :

$$\hat{\kappa}_4(t) = \bar{\kappa}_4(t) - \frac{k_2}{k_4}, \quad (5)$$

where $\bar{\kappa}_4(t) \equiv \hat{\mu}_4(t) / \hat{\mu}_2^2(t)$ is a heuristic estimator which we correct in (5) by suppressing the offset.

We choose the same window for all the involved estimators (i.e., we set $a_4 = a_2 = a_1$), so that

$$\bar{\kappa}_4(t) = \frac{a_1 \mu_4(t-1) + C_1 (x(t) - \hat{\mu}_1(t-1))^4}{(a_1 \mu_2(t-1) + C_1 (x(t) - \hat{\mu}_1(t-1))^2)^4}.$$

For sufficiently large duration of the window ($a_1 \rightarrow 1$), we can treat C_1 as an epsilon and approximate this ratio to the first two terms of its Taylor expansion (this idea was inspired by [4]), which leads to the following expression :

$$\begin{aligned} \bar{\kappa}_4(t) &= (1 + C_1 - 2C_1 \delta^2 x(t)) \bar{\kappa}_4(t-1) \\ &\quad + C_1 \delta^4 x(t) + O(C_1^2), \end{aligned} \quad (6)$$

where $\delta^2 x(t) \equiv (x(t) - \hat{\mu}_1(t-1))^2 / \hat{\mu}_2(t-1)$ computes a normalized distance of the current signal sample to the mean. The correct estimate of the kurtosis proposed in (5) is obtained by subtracting the offset given by

$$\frac{k_2}{k_4} = \frac{-3C_1(4 - 5C_1 + 2C_1^2)}{(2 - C_1)(2 - 3C_1 + 2C_1^2)} = -3C_1 + O(C_1^2),$$

to the approximation in (6). Note that since $C_1 \ll 1$, the offset can be neglected in most practical situations because $\kappa_4 \gtrsim 1$.

The recursive estimation of the kurtosis obtained in (6) is intuitively appealing : if we replace the estimators of the mean and variance by their actual values in the definition of $\delta^2 x(t)$, then noting that $\mathbb{E}[\delta^2 x(t)] = 1$ and $\mathbb{E}[\delta^4 x(t)] = \kappa_4(t)$, we can write

$$\bar{\kappa}_4(t) \approx a_1 \bar{\kappa}_4(t-1) + C_1 \kappa_4(t). \quad (7)$$

Using the equivalence between (1) and (2), we conclude that eq. (6) may be seen as a local average of the crude estimation of $\hat{\kappa}_4(t)$ made by $\delta^4 x(t)$.

Eq. (6) yields a recursive estimation scheme which is described by the pseudo-code in Tab. 1

3. VALIDATION AND PRACTICAL USE

In this section, we make various numerical checks of the proposed method. In all the simulations, the estimator $\hat{\kappa}_4(t)$ as defined in eqs. (5) and (6) is computed with $C_1 = 2.9912 \times 10^{-3}$. This corresponds to a window duration of about 20 s if a sampling rate $f_s = 50$ Hz is assumed.

Check #1: what is the bias of the estimator? — We answer this question in the case of a Gaussian noise for which $\kappa_4 = 3$. Figure 1 (top) shows the histogram of $\hat{\kappa}_4(t)$ computed with a simulated (zero-mean, unit variance and white) Gaussian noise. With this histogram, we evaluate the expectation of $\hat{\kappa}_4(t)$ to be equal to 3.0006 with a bin size of ± 0.03 .

Check #2: is the estimator useful for detecting noises with heavy tails? — Figure 2 presents the result of the estimation of the kurtosis of an evolving mixture of Gaussian and Laplacian noises. The kurtosis of a Laplace random variable is equal to $\kappa_4 = 6$ indicating that the tails of this distribution $\propto \exp(-\sqrt{2}|x|)$ [2] decrease slowly as compared to the Gaussian ones. The two noises are linearly combined ; the weight coefficient of the Laplacian noise increases from 0 to 1 (and reverse for the Gaussian noise) following a linear function of time. An excess of kurtosis (i.e., $\hat{\kappa}_4(t) \gtrsim 4$) appears starting from $t \approx 80$ s (time at which the Laplacian noise starts to dominate) showing that we can answer positively to the question.

Check #3: effect of non-stationarities — Figure 3 illustrates how the recursive estimator of the kurtosis performs with a simulated Gaussian noise of changing mean and variance. After a transient period roughly equal to the window length, $\hat{\kappa}_4(t)$ tends to the correct value which is 3. Each change of the mean or variance is seen as a non-Gaussianity (i.e., large values of the kurtosis). The reason is that the hypothesis of local stationarity required for a correct estimation (see Sect. 1) is not satisfied at the discontinuity points.

Practical use with gravitational wave data — For technical reasons related to the common data format used by the gravitational wave detectors, it is convenient to fix the output rate of the monitoring indicators to one sample per (GPS) second of data (also referred to as *frame*). This applies to the normality index we would like to set up.

Since the variance of $\hat{\kappa}_4(t)$ is difficult to obtain, we cannot compute a confidence interval which would be required to conclude on the normality of the data from the estimator value. We remedy this with the following scheme :

1. choose arbitrarily a threshold η (e.g., $\eta = 4$),
2. associate a warning flag to each frame of data where an excess of kurtosis (i.e., $\hat{\kappa}_4 > \eta$) has been observed at least once,

3. compute the rate of triggered frames (over periods corresponding to the typical duration of a GW as seen by the detector),
4. compare this rate to the one evaluated with Monte-Carlo simulations using a Gaussian noise with similar spectral characteristics than the signal being observed.

Figure 1 (bottom) gives the expected value (with error bars) of this rate if the signal is Gaussian and white, for thresholds taken between 2.4 and 4. For instance, if we fix $\eta = 4$, then rates of triggered frames larger than 1% indicate the presence of a heavy-tailed noise in the data.

```

choose a window duration -> define a1
C1=1-a1
C2=(1-a1^2)/2
bias=-3 C_1

init mul_last, mu2_last, k4_bar_last

while (data available)
  x = get next data sample

  mul= a1 mul_last + C1 x
  dx2= (x-mul)^2
  mu2= a1 mu2_last + C2 dx2
  dx2= dx2/mu2_last
  k4_bar= (1+C1-2 C1 dx2) k4_bar_last ...
          ... + C1 dx2^2
  kappa4= k4_bar + bias

  send kappa4 to output

  mul_last= mul
  mu2_last= mu2
  k4_bar_last=k4_bar
end

```

Table 1. Pseudo-code for the computation of $\hat{\kappa}_4(t)$. The algorithm requires a total number of 16 floating point operations (10 multiplications and 6 additions) to compute the next kurtosis value from the previous one. The memory usage is restricted to three registers of real numbers. As a comparison, a sliding estimate using k-Statistics [5] would need at least a register of the same size than the window. Note that there are several possible initializations of the three registers in line 5. In our simulations, we set them to 0, 1 and 0 respectively. This initialization affects essentially the transient period at the beginning of a computation.

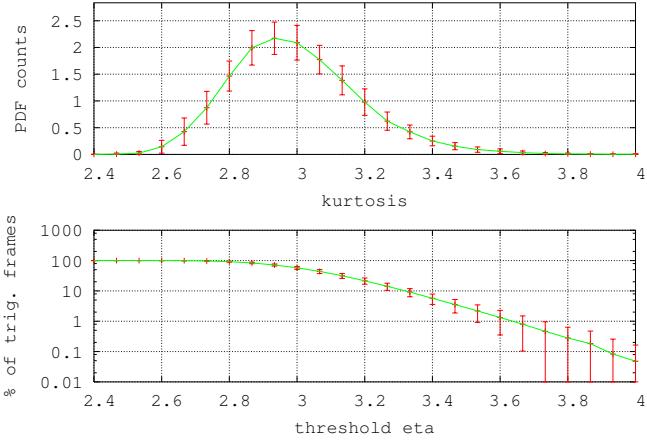


Fig. 1. Probability density function of $\hat{\kappa}_4$ and rate of triggered frames computed with Gaussian noise [6]. The results presented in this figure were obtained using 50 streams of simulated (zero mean, unit variance and white) Gaussian noise. Each data stream contains 30,000 samples (i.e., 600s if the sampling rate is $f_s = 50$ Hz). *top*: the empirical histogram presented in this diagram gives an estimation (bounded by error bars) of the PDF of $\hat{\kappa}_4(t)$ (see Sect. 3 for details). *bottom*: for a threshold η taking values between 2.4 and 4, we present the percentage of frames of data where $\hat{\kappa}_4(t) > \eta$ at least once. To make this computation, the data was divided into 600 chunks of 1 second duration (each of them defining a frame) and the first 20 frames (i.e., equivalent to the window duration) of each trial were removed.

4. REFERENCES

- [1] ,” Here is a list of Internet sites where more information can be found on respective detectors : LIGO (<http://www.ligo.caltech.edu>), TAMA300 (<http://tamago.mtk.nao.ac.jp>), GEO600 (<http://www.geo600.uni-hannover.de>), VIRGO (<http://www.virgo.infn.it>).
- [2] N. L. Johnson and S. Kotz, *Distribution in Statistics. Discrete distributions – Continuous Univariate distributions*, Wiley, New York, 1970.
- [3] P. O. Amblard and J. M. Brossier, “Adaptive estimation of the fourth-order cumulant of a white stochastic process,” *Signal Processing*, vol. 42, no. 1, pp. 37–43, 1995.
- [4] R.M. Aarts, R. Irwan, and A.J.E.M. Janssen, “Efficient tracking of the cross-correlation coefficient,” *IEEE Trans. on Speech and Audio Proc.*, vol. 10, no. 6, pp. 391–402, 2002.
- [5] J. F. Kenney and E. S. Keeping, *Mathematics of Statistics*, Van Nostrand, New York, 2nd edition, 1951.
- [6] ,” All simulations and figures shown in this article were made with GNU Octave and Gnuplot. <http://www.octave.org>.

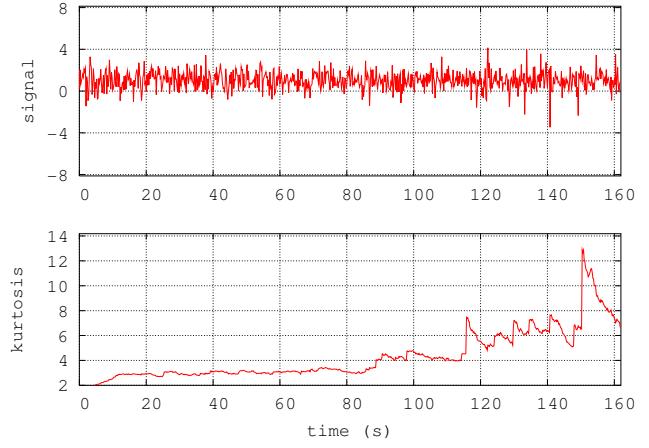


Fig. 2. Recursive estimate of the kurtosis of an evolving mixture of Gaussian and Laplacian noises [6]. *top*: the test signal we use is a mixture of noises (both of unit mean and variance) having resp. a normal PDF $\mathcal{N}(1, 1)$ and a Laplace PDF $\propto \exp(-\sqrt{2}|x - 1|)$. The two noises are combined linearly. The weight coefficient of the Laplacian noise increases from 0 to 1 (and reverse for Gaussian) following a linear function of time. *bottom*: we see in this plot that an excess of kurtosis $\hat{\kappa}_4(t)$ (see Sect. 3 for details) appears starting from $t \approx 80$ s (time at which the Laplacian noise starts to dominate).

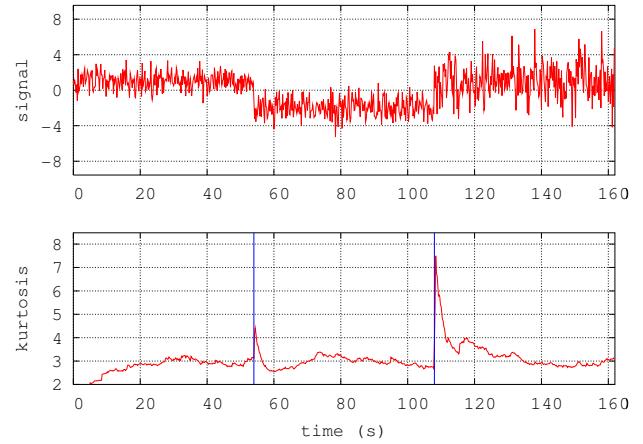


Fig. 3. Recursive estimate of the kurtosis of a Gaussian noise of changing mean and variance [6]. *top*: this is the same simulation as in Fig. 2 with a white Gaussian noise whose mean and variance are set resp. to the following values $(1, -2, 1)$ and $(1, 1, 4)$ during the three periods ($t < 54$ s, $54 \leq t < 108$ s, $t > 108$ s). *bottom*: after a transient period roughly equal to the window length, $\hat{\kappa}_4(t)$ tends to the correct value which is 3. Each change of the mean or variance is seen as a non-Gaussianity (i.e., large values of the kurtosis).